DOCUMENT RESUME

ED 248 207 SP 024 994

AUTHOR Bulcock, Jeffrey W.
TITLE Why Can't We Define Good Teaching?
PUB DATE May 84
NOTE 41p.; Paper presented at the Annual Meeting of the Canadian Society for the Study of Education (Guelph, Ontario, June 1984).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classroom Observation Techniques; College Faculty; *Evaluation Methods; Evaluation Utilization; Foreign Countries; *Formative Evaluation; Higher Education; *Summative Evaluation; Teacher Administrator Relationship; Teacher Behavior; *Teacher Effectiveness; *Teacher Evaluation; Teacher Improvement; Teaching Skills
IDENTIFIERS *Canada

ABSTRACT

Two methods of teacher evaluation, summative evaluation and formative evaluation, are examined. The summative theory calls for an assessment of teaching competence on the basis of which judgments may be made on matters pertaining to promotion, tenure, merit salary increments, or dismissal. Summative evaluation is strongly supported because it provides protection against arbitrary decision making about personnel. However, there are drawbacks to this evaluation theory. Without a clearly articulated theory of teaching, there are many possibilities for errors of judgment on the part of the evaluators. Another problematic feature of summative evaluation is seen to lie with its bureaucratic consequences. Formative evaluation calls for the improvement of teaching through self-identification of a teacher's strengths and weaknesses through interaction and feedback. At each step of the evaluation process, the teacher improves performance, based on regular feedback which increases teacher effectiveness. In this approach, care is taken to avoid ranking the teacher or making judgments about the teacher's ability. The merits and drawbacks of each of these evaluation methods are discussed, and practical implications are considered. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. W. Bulcock

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED248207

WHY CAN'T WE DEFINE GOOD TEACHING?

Jeffrey W. Bulcock
Memorial University of Newfoundland

May 1984

Paper prepared for presentation at the Annual Meetings of
the Canadian Society for the Study of Education
(Canadian Association of the Foundations of Education)
Guelph, Ontario
June 1984

WHY CAN'T WE DEFINE GOOD TEACHING?

1.0 Introduction

The revised terms and conditions of employment for faculty at Memorial University directs department heads to "conduct an annual evaluation of each faculty member under his/her administrative jurisdiction ..." One of the most important criteria to be taken into consideration by the head is teaching. This requirement is in keeping with the charter of the University which enjoins the University to give instruction in every branch of knowledge and learning. A close reading of the University Constitution leads one to conclude that in terms of function the University has to place teaching on an equal footing with research.

Insert Charts 1 and 2 about here

The collective bargaining agreement between the Government of Newfoundland, the Federation of School Boards of Newfoundland and the Newfoundland Teachers Association also recognize the right of school boards to conduct teacher evaluations.

Insert Chart 3 about here

Similar requirements which may be more (or, occasionally less) explicit concerning teacher evaluation; whether at the university/college level or at the public school level, will be found in the legislation, orders in council, terms and conditions of employment or collective bargaining agreements in every province of Canada. Here, we shall focus on university teaching, but in general, the arguments used in addressing the question: "Why can't we define good teaching?" hold for teaching in elementary and high schools as well.

The concept of evaluation may be defined as follows: to determine the significance or worth of something, usually by careful appraisal and study. In order to evaluate something the evaluator normally has to be able to define it, in the sense of being able either (i) to identify its essential equalities, or (ii) to set forth its meaning.

Insert Chart 4 about here

While it is not problematic to set forth the meaning of teaching (for example, teaching involves instruction by precept, example or experience through acting or practicing the profession of a teacher), the thesis held here is that, to date, teaching defies clear definition in the sense that its essential qualities have neither been identified nor described.

Insert Chart 5 about here

We examine this thesis in the remainder of the paper. The argument is simple: How can one evaluate something which has not been defined or described?

2.0 Why Evaluate Teaching?

Four reasons are often given for evaluating teaching performance:

(i) to give administrators information as a basis for making tenure, promotion and merit increment decisions;

(ii) to give individual teachers feedback on their teaching performance primarily as an aid for improving their teaching in the future;

(iii) to provide published information which students may use in course selection;

(iv) to provide research data for institutional research purposes and for an individual researcher's research project.

The reasons are not exhaustive. There are several variations on some of these themes of which three are mentioned:

(v) to protect students from incompetent teachers;

(vi) to protect the teacher from arbitrary administrative decision making; and

(vii) to reward superior teaching performance through the public identification of outstanding teaching.

Here we shall only address the first two reasons. Each is identified as a separate theory: the summative theory (reason (i) above) and the formative theory (reason (ii) above) of the evaluation of teaching effectiveness respectively. Theory one, the summative function of teaching effectiveness, calls for an assessment of teaching competence on the basis of which judgments may be made on matters pertaining to promotion, tenure, merit salary increments -- even dismissal. Theory two, the formative function, on the other hand, calls for the improvement of teaching through the self identification of one's teaching strengths and weaknesses. In this approach, as opposed to the summative approach, care is taken to avoid ranking the teacher or making judgments about the teacher's teaching ability.

When the terms and conditions of employment call for "an annual evaluation of a faculty member" or when a collective bargaining agreement calls for evaluation files which may "only be viewed by the school board, professional board staff, or school administration" the purpose of the evaluation is usually summative. This type of evaluation is much more controversial than formative

evaluation. The distinction between formative and summative theories of evaluation, nevertheless, are more analytical than empirical; which means that the two, in practice, may be difficult to separate. This difficulty may account for the scepticism, even cynicism, on the part of teachers toward practices designed to evaluate their teaching.

Insert Chart 6 about here

3.0 Summative Theory

3.1 The Case for Summative Evaluation

Policy statements found in the terms and conditions of employment and in collective bargaining agreements are increasingly calling for some form of mandatory evaluation. The question is not "Why evaluate?" or "What should be evaluated?" but, rather, "How?". There is little doubt that the evaluation of teaching is here to stay.

Schools, colleges and universities take their teaching functions seriously. Their officials usually have a mandate for monitoring instruction; consequently, they can be held accountable for the quality of the teaching in their departments and faculties. Most universities place teaching on the same formal footing as research and scholarly publication. University authorities recognize the public relations value of promoting excellence in teaching, and the role that teaching plays in generating public confidence in the institution.

A third set of arguments supporting the summative evaluation of teaching focuses on the importance of the incentives for good teaching. Teaching evaluation data which is used for the granting of tenure and the awarding of promotion and other academic honours provide the incentive for faculty to strive for excellence in

teaching. It is unrealistic, claim its advocates, to expect excellence in teaching in the absence of powerful incentives such as salary and merit increments, promotion and tenure.

Neither should one overlook the importance of economic arguments. In periods of economic restraint and recession, tuition costs often have to be raised in order to meet minimum operating expenditures. Students expect more in terms of service, including quality of instruction and tutorial assistance, in return for their increased investment. Concomitantly, cutbacks in operating revenues have pressured university administrations into the adoption of selective tenuring practices and tight promotion policies.

Insert Chart 7 about here

As the competition for increasingly scarce rewards intensifies, personnel decision making is made more difficult. Personnel decisions have to be justified; have to be made in accordance with due process; and have to be shown to be complete, valid and consistent.

Murray (1980) posits a fifth advantage of summative evaluation; namely, its compensatory function. Too often in the past tenure and promotion decisions have been governed by research and publication. An absence of objective information on teaching has resulted in this criteria being ignored. The summative evaluation of teaching, however, makes it possible for a faculty member to gain tenure and promotion through performance in two areas -- teaching as well as research. Murray claims that most faculty prefer an evaluation system where modest performance in one area may be compensated by good performance in another.

A final argument in favour of summative evaluation is drawn from Kronk and Shipka's (1981) National Education Association Handbook for Faculty Leaders. The authors note that university faculty have always competed for scarce rewards; hence, the

legitimacy of recurrent evaluation. When faculty are not evaluated on a scheduled basis, the door may be left open "to administrative decision making which need not reflect systematically gathered data on faculty performance". Systematic evaluation, therefore, restricts "managerial prerogatives" and may deter an administrator from a course of action which might not be justified in the event of an appeal. On these grounds, the recommendation in the NEA Handbook is that "the best approach is to require evaluation of all faculty on a scheduled basis".

3.2 The Case Against Summative Evaluation

3.2.1 The Means/Ends Correspondence Dilemma. It would seem desirable for the advocates of the formal evaluation of teaching to show that there is a significant degree of association between a teaching evaluation and subsequent improvement in teaching performance; in other words, that teaching evaluations result in improved teaching. There is no such evidence. By the same token, it would seem desirable for the advocates of summative evaluation to demonstrate unambiguously that positive teaching evaluations are associated with positive changes in student learning. Again, to date, there is no such evidence.

The absence of findings demonstrating a strong positive relationship between teaching evaluations and student learning is odd. Here we make two observations. In the first place, efforts to measure teaching effectiveness on the basis of its impact on student learning are unpopular. This is because most teachers recognize that student learning is a multicausal activity; and that many significant factors affecting student learning -- family background, motivation, aptitude, course load, class size and organic considerations -- fall well outside the control of the teacher. Thus, to hold teachers responsible for the learning behaviours of their students is unreasonable.

In the second place, there would seem to be a tacit understanding of their dilemma on the part of the evaluators. In one study Gosling (1978) found that over eighty percent of the comments that administrators made on rating forms were positive; and only six percent were of the "needs improvement" variety. Gosling makes the point that administrators account for the observed weaknesses of the students in the classes being evaluated in terms of factors external to the classroom and institutional setting.

3.2.2 Theoretical Naivete 1: Dimensions of Teaching. Theories of teaching are underdeveloped. Few studies, for example, provide a theoretical rationale for selecting the dimensions of teaching which are included in teacher evaluation rating scales. Instead, a host of political, pragmatic and parochial factors are invoked when deciding what dimensions of teaching to include and which to exclude.

When the writer began school teaching in the mid-1950s, teaching for democracy and promotion of the democratic way of life was a "hot" issue. By the 1970s the theme had been abandoned. In the 1960s, student participation in university government was an issue. Faculty members in the arts and sciences in particular were encouraged to consult with their students about grades, examinations and course content. Today, few teacher rating forms refer to the principle of student participation. In short, there is little theoretical basis to the teacher evaluation instruments used today. They tend to reflect ephemeral concerns rather then enduring and more invariant matters.

Theories of teaching do exist. On the basis of social learning theory, one could identify the essential dimensions of the classroom environment: for example, the stimulus or opportunity to learn dimension, the reinforcement or reward/punishment dimension, the expectation/encouragement dimension, and the role model dimension.

Structural-functional theory provides an alternative characterization of teaching: one based on the four functional imperatives inherent in any action system such as a school classroom; namely, adaptation, goal attainment, integration and pattern maintenance (or latency). These have been translated into the domains of the classroom environment by Australian researchers Williams and Batten (1980) as follows: the opportunity domain, the adventure domain, the identity domain and the status domain respectively. The teacher who meets the personal fulfillment expectations of students in these domains is likely to be an effective teacher.

Theories such as social learning theory and structural-functional theory have not been used to formulate models of teaching. Thus, current research efforts tend to be largely atheoretical. Of course, the two theories which were identified may be complementary theories, but given the current state of the teacher evaluation research whether they are or not remains unknown.

3.2.3 Theoretical Naivete 2: The Problem of the Holistic Approach. There have been numerous studies which have examined the relationship between teacher ratings and student achievement. The correlations in these studies have varied widely -- between +.75 and -.80. Abrami et al. (1982:458) account for these inconsistent findings on the grounds that different investigators use different research designs. Some designs are flawed, but several studies such as those by Centra (1977), Ellis and Rickard (1977). Shavelson et al. (1976) and Sullivan and Skanes (1974) have been identified as having strong designs. In each of these strong design studies, however, the assessment of teaching performance was based on a single question on the rating scale. This item calls for the evaluator to make an overall assessment of teaching ability in terms of an order of "goodness" from most desirable to least desirable. The usual labels include: outstanding, excellent, above average, below average, needs improvement, poor or some

variant thereof. The problem is that this type of global item does not represent a dimension of teaching. Another snag is that holistic assessments, because they are based on a single item, are likely to be biased. But the reliability of single item variables of this kind cannot be calculated; thus, it is not possible to estimate their bias. And if the bias cannot be estimated it is not possible to correct for bias.

Researchers using holistic measures of teaching effectiveness are usually able to show that they correlate more highly with student learning than do specific ratings on the various dimensions of teaching behaviour. And this may account for the reason why holistic evaluations are the ones most frequently used for making critical personnel decisions about promotion and tenure. The researchers, however, are well aware of the undesirable psychometric properties of their instruments; thus, most like Sullivan of Memorial University (in a faculty seminar) recommend that holistic variables be used solely in formative evaluations.

Insert Chart 8 about here

Formative evaluation, however, calls for the improvement of teaching through feedback processes whereby a teacher's strengths and weaknesses are identified. It is hard to say, therefore, what use an overall, or holistic, assessment of teaching would be in this process. A holistic assessment does not represent a component or dimension of teaching; and because such information does not relate to the particular strengths or weaknesses of teaching, from the point of view of formative evaluation, the information is gratuitous -- at best it is redundant.

3.2.4 Theoretical Naivete 3: The Problem of Generalizability. Models of teaching effectiveness are sometimes seriously misspecified because important structural components are missing or because structural components were included which should not have been, or

because the model was functionally or mathematically misspecified. One specification problem concerns generalizability. The dimensions of teaching which can usually be found on teacher evaluation rating instruments usually include the following: organization, discipline, atmosphere for learning, sense of humour, health and vitality, appearance, enthusiasm, handling routine procedures, and communication skill. Appropriate distractors such as strongly agree through strongly disagree, or outstanding through needs improvement, are attached to each dimension. It is assumed, therefore, that each dimension is an observable phenomenon. Theory at this level of analysis is concerned with the explanation of relationships between sensory date or observables.

The analyst wishes to know two things: first the magnitude of the relationship between each dimension of teaching and overall teaching performance (teaching style); and second, the magnitude of the relationship between these observables and student achievement. According to the theory a well organized teacher, in control of a well disciplined class, with a sense of humour and with the ability to communicate his enthusiasm for the subject will be given a high rating by the evaluator(s). By the same token, the highly rated teacher along these dimensions is likely to stimulate high student achievement, other things equal.

But what about lecture planning, clarity of lecture presentation, rapport with students, use of instructional aids (the blackboard, models, video tapes), punctuality, and instructional skill? The number of possible dimensions one could add to the list of dimensions could go on and on. In the absence of an articulated theory of teaching there would seem to be no framework for governing the list of potential dimensions. We have noted that both social learning theory and structural-functional theory provide such a framework but, to date, both remain unused by researchers.

3.2.5 Theoretical Naivete 4: Practical Implications. Explanation in the form of observables has several advantages. First, such explanations are not far removed from the conventional wisdom. As such they are intuitively appealing. Second, empirical explanation is usually easy to test; especially at the bivariate level. Third, empirical explanation is practical in the sense that the practical implications usually flow directly from the findings themselves.

In numerous studies of teacher effectiveness, the enalysts have emphasized three findings. First, the correlations between the dimensions of teaching (observables) and student outcomes in terms of achievement are usually modest. Second, the correlations between the dimensions of teaching themselves are often substantial; for example, well organized teachers tend to be enthusiastic teachers and so on. Third, while teachers who have received a low summative rating on a particular dimension can usually increase their ratings on the dimension in question, it is seldom that these efforts have any significant effect on their overall global ratings. The teacher, for example, who is rated as mediocre on "examination question format" is usually able to write better tests the next time round. But the improvement on the particular dimension is unlikely to affect a significant change in an overall rating.

In other words, while empirical explanation is easily tested the findings are sometimes not very much help. One solution to the problem calls for the rejection of empirical explanation in terms of observables, in favour of conceptual explanation in terms of latent constructs or theoretical concepts which are unobservable. The principle of generalizability has been invoked to override the principle of inclusiveness or (misplaced) concreteness. The notion is that the purpose of modelling a complex behaviour such as teaching is not to mirror social reality, but rather to reduce that reality to a set of fundamental and more manageable features.

It is to be regretted, therefore, that this approach to the estimation of theoretical models of teaching effectiveness has never been tried.

3.2.6 Theoretical Naivete 5: The Jackson Effect. At a recent faculty seminar on the determinants of faculty effectiveness, Michael Jackson (editor-in-chief of the Canadian Journal of Education) commented that he found that students in one class would respond differently to his instruction than students in another class; even when he was teaching two sections of the same course. According to the standard explanation of teaching effectiveness, Jackson's teaching should have had much the same effect on one class as the next when the classes are identical in that both are given the same lecture and the same course. Note, too, that the university registration procedure is such as to ensure the random distribution of students between course sections.

Jackson, of course, was challenging the standard model specification in which it is assumed that the teacher influences the student but not the other way round. Students are not lumps of clay which can be molded at will. As every practitioner knows, students possess the capacity to influence the instructional environment in a manner which is presumably to their own advantage. The capacity, however, may differ from class to class. If the students look puzzled perhaps another example of the point being made is in order. If the students do not ask questions about a certain lecture topic perhaps they have grasped it and the instructor can proceed.

Students can influence instruction by asking or not asking for clarification and explanation, by participating or not participating in class discussion. The flow of influence in the classroom is not one way from instructor to student, but two way from instruction to student and back again in a reciprocal manner. But the only model tested in the literature is the one way flow of effect model.

Technically this is known as the recursive model. The recursive model, however, does not incorporate the Jackson effect; thus, it is seriously misspecified. The correct specification is a two way reciprocal, or nonrecursive model. The fact remains, however, that in the research literature on teacher effectiveness, two way, nonrecursive models have neither been formulated nor estimated; thus, the Jackson effect has never been investigated.

3.2.7 Theoretical Naivete 6: The Tautological Hypothesis. We have argued above that models of teacher effectiveness are seriously misspecified in a number of ways. A final specification error will be noted; namely, the fact that in several studies of teaching effectiveness there is a tautological component. When students are used as raters of teaching effectiveness, they usually do so toward the end of a course of instruction. By then they will have seen enough of the instructor to be able to make informed judgments. But by the end of a course students will also have received feedback from tests, laboratory reports and quizzes informing them about their course performance. The question arises, then, whether the prior knowledge biases their ratings. Is it likely that students who are performing well in a course will give a weak instructor a low rating? By the same token, will students who are performing poorly in a course give a good instructor a high rating? If not, then the observed relationship may have a tautological component. In fact, it is well established that grade inflation is an effective way of gaining a high student rating. If the student is obtaining high marks the professor must be good!

In order to take the tautology hypothesis into account the researcher should control for student achievement. In the ideal case the students' gain scores (the score at the end of a course minus the score at the beginning of a course) will be used as a covariate. But such a model calls for a longitudinal design, involving data gathering at at least two points in time. The writer knows of no such study in the teacher effectiveness research literature.

3.2.8 The Dr. Fox Effect. Several observers of the teacher evaluation research literature have noted that the very visit of an evaluator to a classroom can alter the teaching. What an evaluator observes may not represent the typical day to day teaching. To demonstrate this susceptibility of observers to variations in teaching behaviours, lecturers were trained to be slick, enthusiastic and dynamic, while at the same time teaching nonsense. In several replications of the experiment these "teachers" were able to generate high marks both from students and professional "peers" who were asked to rate their performance. Such bias in the ratings is known as the "Dr. Fox effect". It may be found in almost any situation in which a teacher is being assessed.

Observer Bias. Scriven (1980) -- a former President of the American Educational Research Association -- argues that the summative evaluation of teaching is not only wrong, but "a disgrace". He points out that every observer brings to the evaluation situation a set of implicit and explicit assumptions about teaching that will influence the interpretation he places on what is observed. In short, the observer influences the process.

Would the applied statistician be comfortable with an evaluator who was a topologist, or "pure" mathematician? Would the clinical psychologist be prepared to accept the pronouncements of a behavioural psychologist about his teaching? Would the ethnomethodologist regard the teaching assessment of a mathematical sociologist as authentic? Why should the grade one reading teacher accept the ratings of a school administrator who has spent most of his teaching career teaching high school level chemistry?

Art Ponder, one of the writer's colleagues at Memorial University, recently conducted some ingenious research designed to identify observer bias. First, he identified the dominant teaching orientation preferences of a large number of school administrators in terms of Broudy's (1973) triad -- philetics, didactics and

heuristics. Next, each administrator was asked to rate three hypothetical teachers in terms of their teaching effectiveness. One teacher was described in terms of a philetic orientation, one in terms of a didactic orientation and the third in terms of a heuristic orientation. Ponder's hypothesis was that if there were observer bias, philetic oriented administrators would rate philetic teachers more highly than didactic or heuristic teachers; similarly, didactic administrators would rate didactic teachers highest, and heuristic administrators would favour heuristic teachers. His hypothesis was strongly confirmed. An interesting serendipitous finding was that administrators were almost evenly divided between the three orientations.

Insert Chart 9 about here

3.2.9 Bureaucratic Effects. The consequences and implications of accepting summative evaluation is seldom invoked as a reason for deciding to reject it, but in the writer's opinion they are worth considering. Educational administrators use summative evaluations of teaching for decision making purposes. During the present period of recession and economic cutbacks they are cognizant of the sensitivity of teachers toward judicial matters and of the litigious spirit permeating faculty associations. This spirit is congruent with the times -- times in which the recently patriated Constitution is being used to solve a host of outstanding legal issues from the rights of minorities to the rights of parents to decide the kind of schooling they wish their children to have.

It is not sufficient for an administrator to discern that a faculty member needs to improve, or that his teaching is weak. It is necessary to be able to show impartially that the grounds for such judgment are valid. Thus, individuals have to be given reasons for adverse decisions. The reasons have to be given promptly

and the individual given sufficient time to correct the noted deficiencies. This means that institutions have to develop programs designed to assist faculty members who wish to improve their teaching performance, even though participation in such programs will be voluntary. Those adversely affected by an evaluation have to have opportunities of responding in writing and have the right to request information regarding the accuracy (reliability), relevance and validity of the evaluation. Appeals procedures have to be established including rules governing the right to representation at appeals hearings. Institutionalizing summative evaluation means that the standards of performance of faculty members have to be made explicit at the time of appointment.

These are only minimum safeguards. If they are necessary they will prove costly. They will also prove demanding in terms of an administrator's time and energy. The summative evaluation procedures have to be monitored by the interested constituencies. The safeguards for due process will be monitored by the faculty effectively creating a strong faculty organization -- an organization which would soon become proficient in the negotiation and enforcement of master contracts between governing bodies and their employees.

Insert Chart 10 about here

4.0 Formative Theory

4.1 What is Formative Theory?

As we have noted in the above discussion of the advantages and disadvantages of the summative evaluation of teaching, the empirical literature is shot through with conflicting findings. Nevertheless, if the literature supports anything it is that teachers can improve their teaching skills if they are given

constructive feedback on their strong and weak points. Constructive feedback is the key to understanding formative theory. An individual begins with some model of what a good teacher does in a classroom. He receives information or feedback about which characteristics of the model of good teaching seem to work for him and which seem not to work. At this point the individual is faced with several choices; namely, which characteristics to keep or promote, and which characteristics to reject or reduce on the basis of whether they worked or not. Further feedback promotes additional corrections. At each step the teacher becomes a better teacher. The greater the opportunity the individual has for practicing his teaching style coupled with regular feedback the more effective the teacher will become.

It will be the new teacher who is likely to make the greatest incremental improvements in teaching. This is because it is the new teacher who is the most likely to be developing his teaching style. The recurrent and patterned feedback process is characteristic of the formative model. In the ideal case it will result in the adoption of a style of teaching which is the most effective for the individual given his personality resources. The theory, then, gives credence to the adage that "you can't teach an old dog new tricks". Because the feedback process is iterative and incremental, experienced teachers are much more likely than beginning teachers to have adopted a style which is successful for them.

4.2 Formative Theory 1: Selected Observations

4.2.1 On Teacher Characteristics. Formative theory is based on an individual's preliminary conception of a model of good teaching; hence, on notions of the characteristics of good teaching. The model is obviously constrained by the number of characteristics that are available to the individual. By "characteristics" one is

referring to the components of the individual's personality -- physique, appearance speech, gestures, age, sex, humour and the like. These qualities differ from individual to individual. Thus, what works for one teacher need not work for another. Some teachers, for example, are comfortable in an informal discussion-like class situation while others perform better in a more formal lecture-type situation.

4.2.2 On Teacher Effectiveness. The outcome measure in the formative model of teacher evaluation is improved teacher effectiveness. Just as different individuals have different characteristics available to them, so they have different conceptions of what is meant by effective teaching. In that effective teaching is the goal, then different conceptions of the goal will necessitate the use of different means for attaining it. In terms of Broudy's (1973) modes of teaching, for example, one would assume that the mode would be governed by the teaching environment and that the teaching processes underlying each mode would be different from one another.

Teachers, then, have to match their personality characteristics to their notions of effective teaching. In turn, this means that they have to weigh and evaluate their personality attributes in the course of adopting an optimum teaching style, or in the course of maximizing their teaching effectiveness. In general, the teacher's selection of a style or mode of teaching will stimulate the formative process. The more effective the teacher's use of his personality resources for achieving a desired style, the greater the probability of obtaining positive student evaluations. Even so, there can be no guarantee that a positive evaluation will be synonymous with teacher effectiveness, where by teaching effectiveness one is referring to the maximization of student learning. The definition of effective teaching which results in the maximization of an evaluator's assessment evaluation of teaching performance is not one that will necessarily guarantee the maximization of student learning (Lima, 1981:1056).

4.2.3 On Teacher Utility. The maximization of teaching effectiveness via a formative model of teaching effectiveness is still only a subsidiary component of the overall problem; namely, that of maximizing professor utility. The utility function only depends in part on teaching effectiveness. In addition factors such as research, administration and committee work have to be taken into consideration.

Recently, John Centra, who is arguably the foremost authority on the evaluation of teaching performance has listed those factors that are most commonly used in evaluating faculty performance (the utility function). The seven most important factors in order of importance are: classroom teaching, quality of publications, number of publications, personal qualifications (honours, awards, degrees, etc.) research, supervision of graduate student research and committee work. In relation to these factors, activities in professional societies, student advising, community service, consulting and competing job offers were considered to be either minor factors or of negligible importance.

But these utility functions will differ from individual to individual just as teacher characteristics and perceptions of teacher affectiveness do. The individual who places emphasis on the teaching factor is likely to adopt a different style of teaching than the professor who places emphasis on research. In view of this possibility it becomes important that faculty members be aware of the weighting schemes used by tenure and promotion committees because these may have a bearing on the type of style adopted. It is not unreasonable to expect an individual's teaching style to be influenced by their expected institutional rewards.

What do these observations mean in terms of formative theory? They all support a model of teaching effectiveness in which different teachers are likely to have different modes of teaching. The formative theory does not represent so much a model of teaching as

a family of models. The characteristics selected in order to optimize effective teaching will differ, therefore, from one individual to the next depending on the style preferred. The style adopted may also depend on the teacher's conceptions of teaching effectiveness. Teaching effectiveness, in turn, may be influenced by the weighting schemes used by tenure committees. It is conceivable, for example, that two individuals with roughly equal personality resources would develop different modes of teaching based on the expected rewards at their respective universities. A strong research orientation would likely promote a heuristic teaching style. A strong teaching orientation might promote a philetic or human relations style. Aggregate teaching styles in graduate schools might differ appreciably from those in four years liberal arts colleges.

4.3 Formative Theory 2: Research Issues

4.3.1 Measuring Change. The formative model is one fraught with technical difficulties. Three will be discussed: namely, measuring change, nonlinearity and measuring utility function constraints.

In terms of formative theory an individual's teaching effectiveness changes over time as a function of constructive feedback. The model is dynamic rather than static. Yet, of the more than two thousand references that Murray (1980) located on teacher evaluation over the past fifteen years only a handful -- less than one percent -- have been studies of the dynamics of teaching effectiveness. Researchers in education have found it more convenient to hypothetically stop the moving system in order to examine its cross-sectional features rather than look at time-series data which would permit the researcher to examine changes in the magnitude of variables over time. Effectively this means that formative models of faculty teaching effectiveness have not been formulated,

and neither have they been estimated. We have developed our models of teaching by looking at static models when we should have been looking at dynamic models instead.

4.3.2 Nonlinearity. Consider a personality characteristic such as "sense of humour" which in theory should contribute positively to effective teaching. Humour, however, will probably only contribute positively up to a certain level. Beyond a certain point humour might detract from effective teaching. While jokes up to a point will contribute positively to effective teaching, the teacher who spent most of his class time telling jokes is unlikely to enhance his teaching effectiveness. Similarly, the teacher who places excessive dependence on only positive reinforcement or who sets unrealistically high expectations may find his teaching strategies backfiring. These are teaching strategies which are likely to have nonlinear relationships to the preferred mode of teaching. Yet, no model of teaching effectiveness has yet incorporated the polynomial terms necessary to test a nonlinear model.

4.3.3 Utility Function Constraints. As noted above utility functions will differ from individual to individual. If one professor emphasizes teaching and another publication, they are likely to adopt different teaching styles. Yet, seldom is the teaching orientation or any other single orientation an exclusive criterion for career progress decision making. Rogers (1982) has noted that "typically, in making these decisions greater and equal weight is given to teaching and research, and less to service (e.g., 2:2:1 for teaching, research and service respectively)". Time spent on teaching will be constrained by time spent on other factors in the utility function equation. The factors, however, will not be completely independent of one another even though linear independence is a fundamental assumption underlying the accurate estimation of linear equations. Their degree of departure from orthogonality or linear independence will determine the extent to

which the estimates and standard errors will be inflated, and it is difficult to interpret inflated estimates. Though there are ways of coping with these colinearity problems in the teacher evaluation research tradition these remain unknown; hence, unused.

5.0 Summary and Conclusions

5.1 Discussion

Why can't we define good teaching? And if we can't define good teaching how can we evaluate it? The theme of this paper has been that to date teaching defies clear definition. While we can set forth the meaning of teaching we are still unable to identify and describe its essential qualities.

There is a good deal of support for the summative evaluation of teaching. As was noted in section 3.0, the summative evaluation of teaching is here to stay. University charters enjoin faculty members to take their teaching functions as seriously as their research functions. Faculty associations point out that the advantage of summative evaluation is that it provides protection against arbitrary personnel decision making.

Thus, in theory, it is not difficult to support summative theory. But in practice, the implementation of summative evaluation is fraught with difficulties. For example, there is little firm evidence that good teaching necessarily promotes student achievement. This seems to be due to two factors: one is that the variance in achievement is attributable to factors other than teaching; the other is that few well designed studies of the relationship between teaching and learning have been conducted.

There does not seem to be a clearly articulated theory of teaching, for example, as indicated by the absence of a rationale for the selection of the dimensions of teaching included in current teacher rating scales. What models have been formulated and

estimated are often seriously misspecified. They focus on observables (rather than on theoretical constructs); hence, contain far too many dimensions of teaching. These dimensions though highly correlated are but modestly related to measures of teaching effectiveness. Thus, the models lack orthogonality, are subject to sampling fluctuation, and account for only modest variance in either teacher effectiveness or student achievement.

The models also fail to incorporate the necessary feedback relations between teacher behaviours and student classroom behaviours. Prior knowledge of how well a student is performing in a course is seldom controlled for and neither is observer bias. The most problematic feature of summative evaluation, however, lies with its bureaucratic consequences. In order to justify the use of summative methods of teacher evaluation administrators have to be able to guarantee due process. But the very existence of bureaucratic arrangements designed to guarantee due process tend to counteract their own purposes. The machinery can inhibit good teaching because it constitutes a set of constraints on the freedom to teach.

Three conclusions may be drawn from the analysis of formative models of teaching effectiveness. First, and most important, teachers can significantly improve their effectiveness if they are given constructive feedback and the opportunity to develop a teaching style which is congruent with their personality characteristics. Second, formative theory promotes a family of models because different teachers adopt different teaching styles. Third, the formative model is a dynamic rather than a static one. But because of this there are several technical difficulties which have to be coped with if formative theory is to be adequately tested. These difficulties can be overcome if longitudinal, nonrecursive structural equation modelling is used, but to date, the formative model of teaching effectiveness remains an untested theory.

5.2 Practical Implications

(i) Given that the research on teacher evaluation has, to date, been unable to identify or describe the essential qualities of effective teaching and given the inevitability of observer bias there seems little empirical justification for summative evaluations of teaching effectiveness by either colleagues, administrators or students.

(ii) Nevertheless, there are several sound reasons for conducting summative evaluations even though the evaluators may be biased and even though they do not know what constitutes good teaching. If teachers are persuaded by these arguments (see sections 2.0 and 3.0 above) perhaps the disadvantages of the summative evaluation of teaching could be outweighed by their purported advantages; in which case teachers may opt for summative evaluations of their teaching performance.

(iii) Though formative theory has not been adequately tested we are sufficiently encouraged by the promise that the theory holds for improving teaching to recommend that guidelines for conducting formative evaluations of teaching need to be identified and made available (especially to beginning teachers). It is recommended that these instruments include self-appraisal or self-evaluation forms, student rating forms and colleague observation guides. It is also recommended that opportunities be made available for teachers to self-review their videotaped classes.

(iv) Teachers, then, should have the right to conduct recurrent formative evaluations of their teaching. Administrative authorities should be cognizant of the value of formative evaluation early in a teacher's career.

(v) Though teachers should not be required to collect formative evaluation information; nor should they be expected to provide it, if available, to administrative authorities; nevertheless, they should have the right to make the results available to others if

they wish. Teachers exercising this option should be aware that the information is being used in summative fashion.

(vi) Formative evaluation or self-evaluation is likely to be successful only if opportunities are provided teachers both in terms of time and facilities for such evaluations to be conducted. Normally the procedure will involve obtaining student and colleague feedback in non-threatening, non-coercive situations.

References

Abrami, P. C., Leventhal, L. and Perry, R. P. Educational seduction. Review of Educational Research, 1982, 52, 446-464.

Broudy, H. S. The arts of teaching: Didactics, heuristics and philetics. Eighth Distinguished Scholar Lecture, Kent State University, June 1973.

Centra, J. A. Determinants of Faculty Effectiveness. San Francisco: Jossey-Bass, 1980.

Ellis, N. R. and Rickard, H. C. Evaluating the teaching of introductory psychology. Teaching of Psychology, 1977, 4, 128-132.

Gosling, A. W. Teacher evaluation: An examination of formal expectations and actual content. (Doctoral dissertation, Indianna University, 1978). Dissertation Abstracts International, 1979, 39, 5228A.

Kronk, A. K. and Shipka, J. A. Evaluation of Faculty in Higher Education: A Handbook for Faculty Leaders. Washington, D.C.: National Education Association, 1981.

Lima, A. K. An economic model of teaching effectiveness. American Economic Review, 1981, 712, 1056-1059.

Murray, H. G. A Comprehensive Plan for the Evaluation of Teaching at the University of Queensland. The University of Queensland Tertiary Education Institute, 1980.

Rogers, W. T. Canadian educational research: Our responsibility. Canadian Society for the Study of Education News, 1982, IX, 7, 8.1-8.7.

Scriven, M. The Evaluation of College Teaching, In Service. Syracuse, New York: National Council of States of Inservice Education, 1980.

Shavelson, R. and Dempsey-Atwood, N. Generalizability of measures of teaching behavior. Review of Educational Research, 1976, 46, 553-611.

Sullivan, A. M. and Skanes, G. R. Validity of student evaluation of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 1974, 66, 584-590.

Williams, T. and Batten, M. The Quality of School Life. Hawthorn, Australia: The Australian Council for Educational Research, 1981.

CHART 1

CLAUSE 3.2 FROM AN ACT RESPECTING THE MEMORIAL UNIVERSITY OF NEWFOUNDLAND

3(2) THE UNIVERSITY SHALL HAVE FULL POWER AND AUTHORITY FROM TIME TO TIME AND AT ALL TIMES TO ESTABLISH AND MAINTAIN SUCH FACULTIES, COLLEGES, SCHOOLS, INSTITUTIONS, DEPARTMENTS, CHAIRS AND COURSES AS TO THE BOARD OF REGENTS MAY SEEM MEET, AND TO GIVE INSTRUCTIONS AND TRAINING IN ALL BRANCHES OF KNOWLEDGE AND LEARNING ... TO PROVIDE FACILITIES FOR THE PROSECUTION OF ORIGINAL RESEARCH IN EVERY BRANCH OF KNOWLEDGE AND LEARNING AND TO CONDUCT AND CARRY ON SUCH RESEARCH WORK; AND GENERALLY, TO PROMOTE AND CARRY ON THE WORK OF A UNIVERSITY IN ALL ITS BRANCHES.

(EMPHASIS ADDED)

CHART 2

EXCERPTS FROM CLAUSE 8.1 FROM AN ACT RESPECTING THE MEMORIAL UNIVERSITY OF NEWFOUNDLAND

8.1 THE UNIVERSITY SHALL, SO FAR AS, AND TO THE FULL EXTENT WHICH ITS RESOURCES FROM TIME TO TIME PERMIT, PROVIDE

A) <u>SUCH INSTRUCTION IN ALL BRANCHES OF LIBERAL EDUCATION</u> AS MAY ENABLE STUDENTS TO BECOME PROFICIENT IN AND QUALIFY FOR DEGREES, DIPLOMAS, AND CERTIFICATES IN SCIENCE, COMMERCE, ARTS, LITERATURE, LAW, MEDICINE, AND ALL OTHER BRANCHES OF KNOWLEDGE;

B) <u>SUCH INSTRUCTION, WHETHER THEORETICAL,</u> TECHNICAL, ARTISTIC, OR OTHERWISE, AS MAY BE OF SPECIAL SERVICE TO PERSONS ENGAGED ... IN THE FISHERIES, MANUFACTURES OR THE PURSUIT OF ENGINEERING, AGRICULTURAL, AND INDUSTRIAL PURSUITS OF THE PROVINCE.

(EMPHASIS ADDED)

CHART 3

ARTICLE No. 14 - EVALUATION - FROM THE COLLECTIVE BARGAINING AGREEMENT BETWEEN THE TREASURY BOARD, THE FEDERATION OF SCHOOL BOARDS AND THE NEWFOUNDLAND TEACHER'S ASSOCIATION

14.01 - THE PRIME PURPOSE OF EVALUATION SHALL BE THE INCREASED EFFECTIVENESS OF PERSONNEL IN IMPROVING INSTRUCTION.

14.02 - ALL EVALUATION SHALL BE CONDUCTED OPENLY AND WITH KNOWLEDGE OF THE TEACHER(S).

14.03 - THE RESULTS OF SUCH EVALUATION SHALL BE MADE KNOWN TO THE TEACHER(S) CONCERNED AND WHERE RESULTS OF EVALUATIONS ARE PRODUCED IN WRITTEN FORM, A COPY WILL BE GIVEN TO THE TEACHER(S) CONCERNED.

14.04 - PROPER SECURITY SHALL BE MAINTAINED ON TEACHER EVALUATION FILES. THE FILES MAY ONLY BE VIEWED BY THE SCHOOL BOARD, PROFESSIONAL BOARD STAFF, OR SCHOOL ADMINISTRATION. ANY OTHER PERSONS WISHING TO VIEW THE FILES COULD ONLY DO SO WITH THE CONSENT OF THE TEACHER(S). TEACHERS SHALL BE PERMITTED TO VIEW THEIR OWN FILE AT ANY CONVENIENT TIME DURING NORMAL OFFICE HOURS.

CHART 4

Definitions of Terms

1. Evaluation: to determine the significance or worth of something, usually by careful appraisal and study.

2. Define: to determine or identify the essential qualities or meaning of something.

CHART 5(1)

The Meaning of Teaching

"Teaching involves instruction by precept, example or experience through acting or practicing the profession of a teacher."

CHART 5(2)

The Thesis of the Paper on "Why Can't We Define Good Teaching?"

"That, to date, teaching defies clear definition in the sense that its essential qualities have neither been identified nor described."

CHART 6

Some Basic Reasons for Evaluating Teaching

- To give administrators information as a basis for making tenure, promotion and merit increment decisions.
- To give individual teachers feedback on their teaching performance primarily as an aid for improving their teaching in the future.
- To provide published information which students may use in course selection.
- To provide research data on the art of teaching.
- To protest students from incompetent teachers and chartatanism.
- To protect the teacher from arbitrary administrative decision making.
- To reward superior teaching performance.

CHART 7

The Case for Summative Evaluation

- Manditory evaluation of teaching is called for in terms and conditions of employment and in collective bargaining agreements.

- Educational officials have mandates by law for monitoring the quality of instruction.

- Teaching evaluations provide incentives for good teaching.

- Students and their parents pay for high quality instruction.

- Many university teachers welcome summative teaching evaluations because it enables them to gain tenure and promotion through performance in teaching as well as in research and service.

- The evaluation of teaching on a scheduled basis deters arbitrary administrative decision making.

CHART 8.1

Example of a Holistic Assessment

ALL THINGS CONSIDERED, AND IN COMPARISON TO OTHER INSTRUCTORS YOU HAVE HAD IN SECONDARY SCHOOL AND COLLEGE, HOW EFFECTIVE HAS THE INSTRUCTOR BEEN IN THIS COURSE? CIRCLE ONE RESPONSE NUMBER.

ONE OF THE MOST EFFECTIVE (AMONG TOP 10%)	MORE EFFECTIVE THAN MOST (AMONG TOP 30%)	ABOUT AVERAGE	NOT AS EFFECTIVE AS MOST (LOWEST 30%)	ONE OF THE LEAST EFFECTIVE (LOWEST 10%)
5	4	3	2	1

CHART 8.2

Examples of Rating Scale Items

CIRCLE ONE NUMBER WHICH CORRESPONDS MOST CLOSELY TO YOUR OBSERVATION

3 - VERY SATISFIED
2 - SATISFIED
1 - NEEDS IMPROVEMENT
NA - NOT APPLICABLE

- THE MATERIAL PRESENTED IS ACCEPTED BY COLLEAGUES AS WORTH KNOWING	3	2	1	NA
- WHEN APPROPRIATE DIVERGENT VIEWPOINTS WERE PRESENTED	3	2	1	NA
- STATED THE PURPOSE OF THE LECTURE	3	2	1	NA

CONT'D.

CHART 8.2 (CONT'D.)

- EXPLICITLY STATED THE RELATIONSHIPS AMONG VARIOUS IDEAS IN THE LECTURE	3	2	1	NA
- STATED WHAT STUDENTS WERE EXPECTED TO GAIN FROM THE LECTURE MATERIAL	3	2	1	NA
- ASKED QUESTIONS TO SEE WHAT STUDENTS KNEW ABOUT THE LECTURE TOPIC	3	2	1	NA
- GREETED STUDENTS WITH A BIT OF SMALL TALK	3	2	1	NA
- ENCOURAGED STUDENT QUESTIONS	3	2	1	NA
- ANSWERED QUESTIONS SATISFACTORILY	3	2	1	NA

CHART 9

Summary of Specification Errors in Teaching Effectiveness Research

- Identification and description of the theoretically relevant dimensions of teaching
- Limitations of the holistic approach
- The generalizability problem
- The naturalistic fallacy - from "is" to "ought"
- The Jackson Effect - reciprocal or feedback phenomena
- Tautologies
- Observer bias: the question of validity

CHART 10

Some "Due Process" Requirements

1. An individual should receive specific, valid reasons for an adverse decision.

2. An individual should receive prompt notice if his teaching performance is evaluated as deficieny in some respect.

3. Prior to an adverse decision the individual should be given a reasonable length of time to correct the noted deficiency.

4. Institutions using summative methods of evaluation should provide a program designed to assist faculty members who wish to improve their teaching performance.

5. Faculty member participation in such a program should be voluntary.

6. All information on a faculty member's teaching performance should be accessible to him.

7. Faculty members should have the opportunity to respond in writing to an evaluation of his teaching performance; and, in particular, should be able to request information on the accuracy (reliability) relevance, completeness, validity and consistency of the evaluation.

CONT'D.

CHART 10 (CONT'D.)

8. AN INDIVIDUAL SHOULD HAVE AN OPPORTUNITY OF JUDGING ANY INSTITUTIONAL FILE OF PRIOR TEACHING EVALUATIONS AFTER A SPECIFIED TIME AND ESPECIALLY WHEN ALLEGATIONS IN ANY EVALUATION ARE ACCURATE, IRRELEVANT OR INCOMPLETE.

9. ANY EVALUATION WHICH RESULTS IN AN ADVERSE PERSONNEL DECISION SUCH AS NONRENEWAL OF CONTRACT, TENURE DENIAL, PROMOTION DENIAL OR THE WITHHOLDING OF MERIT SALARY INCREMENTS, SHOULD HAVE THE RIGHT TO APPEAL THE DECISION TO AN IMPARTIAL SOURCE.

10. IN SUCH AN APPEAL THE INDIVIDUAL SHOULD HAVE THE RIGHT TO REPRESENTATION.

11. FACULTY MEMBERS SHOULD BE EVALUATED IMPARTIALLY; THAT IS, SHOULD BE EVALUATED IN THE SAME FASHION AS OTHER FACULTY MEMBERS WHO DO THE SAME KIND OF WORK.